Methane-derived carbon flows into host–virus networks at different trophic levels in soil

Sungeun Lee[a], Ella T. Sieradzki[b], Alexa M. Nicolas[c], Robin L. Walker[d], Mary K. Firestone[b,e], Christina Hazard[a,1], and Graeme W. Nicol[a,1,2]

[a]Environmental Microbial Genomics Group, Laboratoire Ampère, École Centrale de Lyon, CNRS UMR 5005, Université de Lyon, Ecully 69134, France; [b]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720; [c]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [d]School of Rural Land Use, Scotland's Rural College, Aberdeen, AB21 9YA, United Kingdom; and [e]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Edited by Caroline S. Harwood, University of Washington, Seattle, WA, and approved June 25, 2021 (received for review March 16, 2021)

The concentration of atmospheric methane (CH_4) continues to increase with microbial communities controlling soil–atmosphere fluxes. While there is substantial knowledge of the diversity and function of prokaryotes regulating CH_4 production and consumption, their active interactions with viruses in soil have not been identified. Metagenomic sequencing of soil microbial communities enables identification of linkages between viruses and hosts. However, this does not determine if these represent current or historical interactions nor whether a virus or host are active. In this study, we identified active interactions between individual host and virus populations in situ by following the transfer of assimilated carbon. Using DNA stable-isotope probing combined with metagenomic analyses, we characterized CH_4-fueled microbial networks in acidic and neutral pH soils, specifically primary and secondary utilizers, together with the recent transfer of CH_4-derived carbon to viruses. A total of 63% of viral contigs from replicated soil incubations contained homologs of genes present in known methylotrophic bacteria. Genomic sequences of ^{13}C-enriched viruses were represented in over one-third of spacers in CRISPR arrays of multiple closely related *Methylocystis* populations and revealed differences in their history of viral interaction. Viruses infecting nonmethanotrophic methylotrophs and heterotrophic predatory bacteria were also identified through the analysis of shared homologous genes, demonstrating that carbon is transferred to a diverse range of viruses associated with CH_4-fueled microbial food networks.

methanotrophy | methylotrophy | viruses | predator | stable isotope probing

Microorganisms play a central role in global carbon (C) biogeochemical cycling in soil systems. Soil is one of the most diverse habitats in the biosphere and can typically contain 10^9 to 10^{10} prokaryotic cells (1) or viruses (2) per g. Infection by viruses facilitates the horizontal transfer of genes, and viral lysis acts as a control of host abundance and releases nutrients. In the marine environment, 20 to 40% of prokaryotes are lysed on a daily basis with the release of 150 Gt of C annually (3). However, the role of viruses in influencing prokaryotic ecology in soil remains comparatively unknown (4). In particular, difficulties remain in identifying the frequency of active interactions between native host and virus populations in situ, largely due to a lack of tools to study interactions within the highly complex and heterogeneous soil environment. While red-queen or "arms race" dynamics have not yet been observed in natural soil populations as they have in marine systems (5), studies have shown viruses can coevolve with their hosts in soil and that hosts change in their susceptibility to infection (6). Shotgun sequencing of diverse soil microbial communities has enabled identification of linkages between viruses and hosts involved in carbon cycling both through identifying CRISPR protospacer sequences in viral genomes and the presence of viral genes encoding enzymes involved in complex carbon degradation (7). However, determining virus–host associations in situ with these methods does not typically elucidate the timeline of multiple viral infections, with linkages potentially associated with populations not active under current conditions or even with relic DNA (8).

The atmospheric concentration of methane (CH_4) has more than doubled since the mid-18th century (9), contributing 25% of additional radiative forcing from persistent greenhouse gases (10). Methanotrophs play a major role in removing atmospheric CH_4, with soils estimated to contribute 5% of the global sink (9) and mediating fluxes to the atmosphere from methanogenic activity in anoxic habitats in soil (11). Consequently, there is considerable knowledge of the diversity and functioning of CH_4-cycling microorganisms in soil (e.g., refs. 12 to 17). Aerobic methanotrophs use CH_4 for both carbon and energy requirements, and key representatives in soil belong to the type I Gammaproteobacteria family *Methylococcaceae*, type II Alphaproteobacteria families *Methylocystaceae* and *Beijerinckiaceae*, and *Methylacidiphilaceae* of the Verrucomicrobia (13). Soil pH is one of many factors influencing methanotroph activity, and type I and type II methanotrophs can dominate activity in neutral and acidic pH soils, respectively (18). In addition, a wide variety of nonmethanotrophic methylotrophs utilize methanol produced and excreted by methanotrophs, and together methanotrophic and other methylotrophic

Significance

The impact of soil viruses on prokaryotic hosts and their functional processes is largely unknown. While metagenomic sequencing of soil microbial communities enables identification of linkages between viruses and hosts, this does not necessarily identify contemporary interactions. To enable a detailed analysis of active virus–host interactions between individual populations, we focused on the critical biogeochemical process of methane (CH_4) oxidation and followed the transfer of carbon from hosts to their associated viruses in situ. Analysis of ^{13}C-enriched metagenomic DNA demonstrated that CH_4-derived carbon is transferred into viral biomass via both primary and secondary utilizers of CH_4 and suggests viral predation is an important mechanism for releasing CH_4-derived organic carbon into the soil matrix.

Author contributions: S.L., C.H., and G.W.N. designed research; S.L. performed research; R.L.W., M.K.F., C.H., and G.W.N. contributed new reagents/analytic tools; S.L., E.T.S., A.M.N., C.H., and G.W.N. analyzed data; and S.L., E.T.S., A.M.N., C.H., and G.W.N. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]C.H. and G.W.N. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: graeme.nicol@ec-lyon.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2105124118/-/DCSupplemental.

Published August 4, 2021.

single-carbon compound (C1)-utilizing consortia assimilate CH_4-derived carbon in a variety of habitats (19).

Despite the substantial amount of work determining the ecophysiology of methanotrophs, very little is known about the role of viruses in influencing their ecology and functioning in any natural environment. Tyutikiv and colleagues (20, 21) characterized the morphology of *Methylosinus*-infecting viruses isolated from a range of habitats including soil, fish, bovine rumens, and terrestrial water sources. More recently, metagenomic analyses have identified viruses predicted to infect methanotroph hosts in

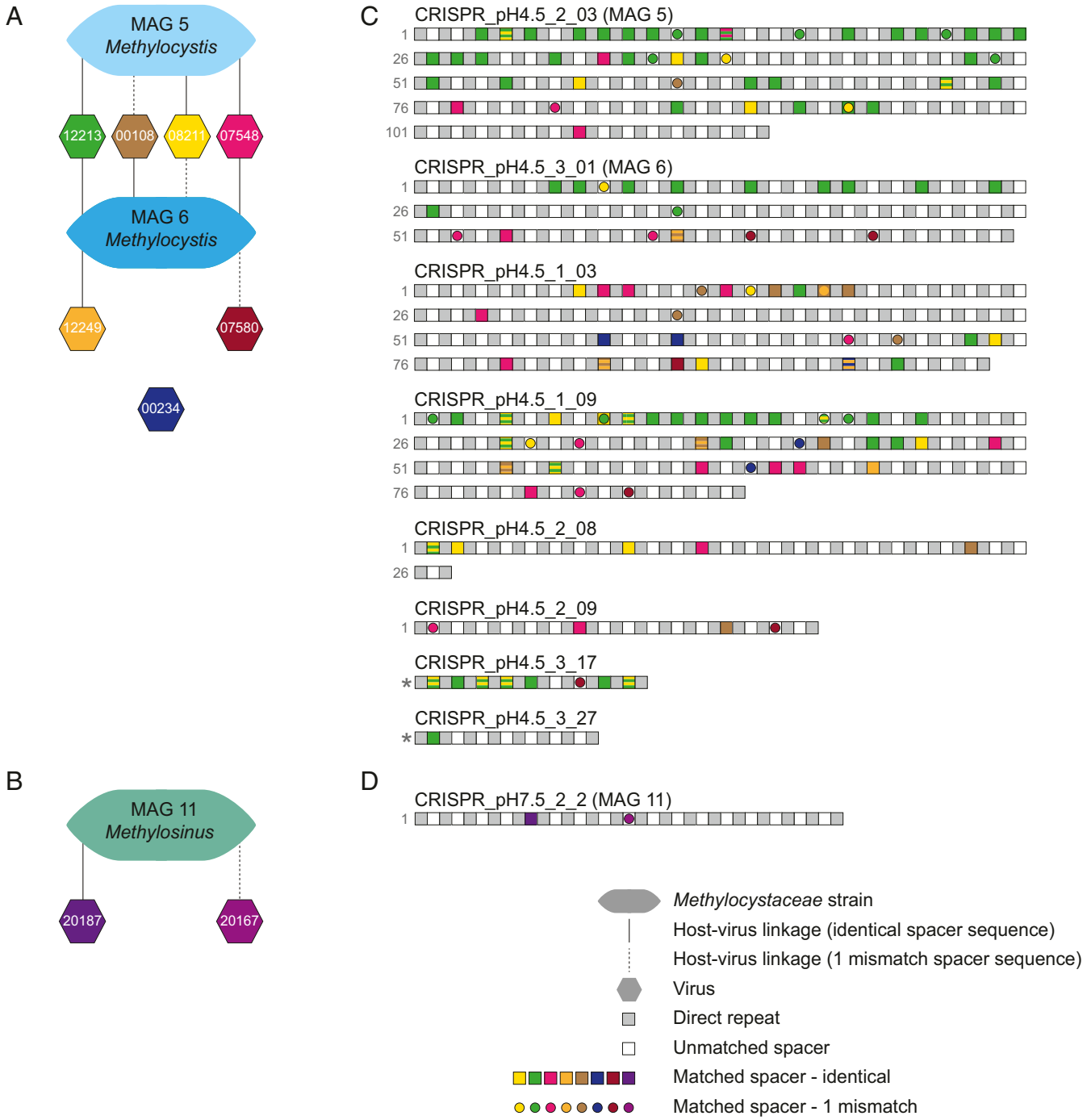

Fig. 1. Active ^{13}C-enriched viruses of individual *Methylocystaceae* populations identified from spacer sequences in CRISPR arrays. (*A* and *B*) Schematic representation of linkages between individual mVCs and *Methylocystis* or *Methylosinus* MAGs, respectively. Shared host spacer/virus protospacer sequences were identified with complete identity or one mismatch. The numbers in hexagons denote mVC IDs, with an unconnected hexagon linked to unbinned CRISPR arrays only. (*C* and *D*) Distribution of spacers from seven mVCs in *Methylocystis* CRISPR arrays (MAGs 5 and 6 and six unbinned contigs) and two mVC in *Methylosinus* MAG 11's CRISPR array, respectively. CRISPR array names describe the individual soil microcosm from which a contig was derived. DRs for complete arrays are numbered (in gray), with the spacer after DR 1 being the most recently incorporated. Two partial arrays are denoted with *. Spacers with complete identity or one mismatch to sequences in mVCs are represented by color-coded squares and circles, respectively, with stripes highlighting sequences found in two different mVCs. Spacer sequences were identified and matched to mVCs using the CRISPR Recognition Tool (37) and Seqkit (38), respectively.

soil (7) and freshwater habitats (22), with the latter study also identifying genes encoding particulate methane monooxygenase subunit C (PmoC) in the genomes of giant viruses, indicating that they may have the ability to augment the activity of infected methanotrophs. Viruses of nonmethanotrophic methylotrophs have also been isolated but from nonsoil environments (23–25). However, we currently have no knowledge of the dynamics of virus interactions with soil methanotroph or nonmethanotrophic methylotroph populations in situ.

A widely used technique for identifying active populations within a diverse microbial community in environmental samples, including methanotrophs, is DNA stable-isotope probing (26). Incorporation of a substrate enriched with an isotope can be traced in genomes of community members. This can demonstrate utilization of the original substrate thereby linking a genome to a functional process but may also be the result of secondary utilization [i.e., incorporation of the isotope from a metabolic product or microbial biomass (27)]. As viruses are entirely composed of elements derived from a host cell, their production inside active hosts incorporating an isotopically enriched substrate will also result in detectable viral isotopic enrichment (28). In this study, we aimed to identify active virus–host interactions within a complex soil habitat by focusing on a taxonomically and functionally restricted group of organisms. By following ^{13}C flow in situ, we aimed specifically to identify DNA viruses actively infecting their methanotroph host, using CH_4-derived C, including the identification of individual virus–host interactions and potentially those actively infecting secondary utilizers such as nonmethanotrophic methylotrophs.

Results and Discussion

Analysis of Methane-Derived ^{13}C-Enriched Virus and Bacterial Metagenomes. After aerobically incubating pH 4.5 and 7.5 soils in the presence of ^{12}C- or ^{13}C-CH_4, high–buoyant density genomic DNA (>1.732 g · ml^{-1}) containing ^{13}C-enriched or ^{12}C-high guanine+cytosine (GC) mol% genomic DNA was recovered via isopycnic centrifugation in CsCl gradients (*SI Appendix*, Fig. S1). Six metagenomes were produced from ^{13}C isotopically enriched DNA samples only (three pH 4.5 and three pH 7.5; Dataset S1). Concentrations of high–buoyant density genomic DNA from ^{12}C-CH_4 incubations were too low for comparable shotgun sequencing. While this indicated minimal recovery of unenriched DNA in ^{13}C-incubated samples, analysis of 16S rRNA gene amplicon libraries prepared from high–buoyant density DNA of both ^{12}C and ^{13}C-CH_4 incubations confirmed ^{13}C-enrichment of C1-utilizing populations (*SI Appendix*, Fig. S2). The most abundant families in ^{12}C amplicon libraries belonged to the phylum Actinobacteria, containing members with high GC mol% genomes whereas the most abundant families in ^{13}C amplicon libraries contained members with known C1 metabolisms (i.e., *Hyphomicrobiaceae* and *Methylococcaceae* in pH 4.5 and 7.5 soils, respectively). This indicates that high–buoyant density ^{13}C-enriched metagenomic libraries represented active C1 incorporators rather than microorganisms with high GC mol% DNA.

Reads from individual metagenomes were assembled before taxonomic assignment of contigs. Reproducibly distinct communities were enriched in the two soils (*SI Appendix*, Fig. S2), with only six bacterial families each representing >1% of reads mapped to contigs ≥5 kb and all including known C1-utilizing taxa (*Beijerinckiaceae*, *Bradyrhizobiaceae*, *Hyphomicrobiaceae*, *Methylococcaceae*, *Methylocystaceae*, and *Methylophilaceae*). We resolved 23 medium- and high-quality (29) metagenome-assembled genomes (MAGs) (Dataset S2), including 12 methanotrophs. Specifically, three MAGs represented gammaproteobacterial type I methanotrophs (*Methylobacter*), and nine MAGs represented alphaproteobacterial type II methanotrophs (*Methylocystis*, *Methylosinus*, or *Methylocapsa*). Secondary utilizers of CH_4-derived organic carbon were also identified, with nine MAGs associated with established or putative nonmethanotrophic methylotrophs, lacking CH_4 oxidation machinery but capable of utilizing methanotroph-derived methanol (*SI Appendix*). These included representatives of the *Gemmatimonadales*, *Hyphomicrobium*, *Herminiimonas*, and *Rudaea*, the latter two, to our knowledge, not having been previously associated with C1 metabolisms but possessed predicted methanol or formate dehydrogenases (Dataset S2). Two MAGs represented strains of *Bdellovibrio* and *Myxococcus*, known predatory bacteria (30), indicating that growing methanotrophic and methylotrophic populations were preyed upon.

Virus populations linked to C1 hosts were analyzed using metagenome viral contigs (mVCs), predicted using established tools. Using contigs ≥10 kb (31), VirSorter (32) predicted 270 mVCs, with a further 4 "likely" mVCs predicted uniquely by DeepVirFinder (33) (*SI Appendix*), together representing 227 viral operational taxonomic units (34). Analysis of the normalized read mapping for mVCs demonstrated that, as with the bacterial communities, active ^{13}C-enriched viral populations were reproducibly distinct between acidic and neutral pH soils (*SI Appendix*, Fig. S3). Analysis of free virus-targeted metagenomes (viromes) prepared from the same soil prior to CH_4 incubation (Dataset S1) revealed that reads could be mapped to 144 (53.3%) of all ^{13}C-derived mVCs, even before enrichment for methane-utilizing consortia. This indicates that the majority of mVCs identified after CH_4 incubation were from free viruses rather than those integrated in host genomes. A total of 34% of mVCs possessed an integrase homolog, a marker gene for a temperate life cycle, which is comparable to the proportion of temperate viruses that constitute free viruses in other environments (e.g., refs. 35 and 36). This suggests mVCs sampled in this study represent a mixture of viruses capable of lysogenic or lytic-only life cycles.

Identification of Active Methanotroph Viruses and Linkage to Individual Populations through Analysis of CRISPR Arrays. CRISPR arrays were identified in 3 of 23 MAGs, each associated with the genus *Methylocystis* or *Methylosinus* of the *Methylocystaceae* (Fig. 1). In the acidic soil, complete CRISPR arrays of growing methanotrophs were associated with two *Methylocystis* MAGs (MAG identifiers 5 and 6) sharing 79.2% average nucleotide identity (ANI) and likely representing different species (39). A further four complete and two incomplete CRISPR arrays were identified in unbinned bacterial contigs all possessing the same direct repeat (DR) sequence. These eight arrays varied in size, ranging from 9 to 114 DRs, contained a total of 432 spacers, and were in the same size range of *Methylocystaceae* CRISPR arrays from sequenced genomes (Dataset S3). Comparison of spacer incorporation between arrays revealed that these multiple closely related populations had different histories of viral interaction. Genome sequences from ^{13}C-enriched viral populations were represented by seven mVCs with 100% sequence identity to 29.5% of spacers. In addition, 7.9% of spacers possessed a one nucleotide mismatch, all of which represented a synonymous substitution, indicating that variation was the result of mutations in viral genomes increasing their ability to evade CRISPR-Cas defense systems or genetic variation in closely related viral populations. Only three pairs of spacers were identical, with each pair member located on a different array. Potential variation in virus host range was also observed, with mVCs linked to only one or both *Methylocystis* MAGs, respectively.

Surprisingly, a large number of spacers in individual *Methylocystis* CRISPR arrays were linked to the same virus, with up to 31 being homologous to protospacer sequences in one mVC. Sequences of adjacent spacers matching the same mVC were verified from individual reads. To provide support that these multiple spacers were derived from *Methylocystis*-associated viruses, mVCs were examined for host-specific conserved protospacer-adjacent motif (PAM) sequences (40). Consistent with the identification of

genuine protospacers, 138 of 146 linked spacers (i.e., all possessing ≤1 mismatch) had the same conserved PAM sequence "TTC" [target-centric orientation (41)]. The relative position of spacers from each mVC in the arrays also revealed temporal differences in virus interaction. For example, sequences from viruses represented by mVC_12213_cat.2 were present in more recently incorporated spacers in some arrays, including the latest integrated spacer in one complete array, revealing the possibility of incorporation during the incubation of the experiment.

Analysis of all CRISPR arrays (i.e., including those in unbinned contigs) revealed that the majority of linked ^{13}C-enriched viruses were associated with methanotrophic populations (Fig. 2*A*). In total, 11 different variants were identified (i.e., each having a unique DR sequence) with 9 linked to *Methylocystaceae* or *Methylococcaceae* populations. DR sequences generally possessed high sequence similarity to those in CRISPR arrays from genomes of cultivated strains of the same family, although only CRISPR array 6 had a DR sequence that was identical (Dataset S3). Individual DR variants were restricted to either pH 4.5 or 7.5 soil. Using 100% sequence identity in searches between CRISPR spacer and mVC protospacer sequences, 19 VirSorter-predicted mVCs were linked to all CRISPR array variants. In addition, analysis of shorter predicted mVCs ranging 5 to 10 kb identified two additional linked mVCs (mVC_08964_cat.3 [9.8 kb] and mVC_28139_DVF [5.1 kb]). One-third of CRISPR-linked mVCs were categorized at the lowest level of confidence [i.e., category-3 by VirSorter (32) or "possible" by DeepVirFinder (33)], suggesting that retaining only higher-confidence contigs may exclude a substantial proportion of bona fide methanotroph virus–derived contigs in uncharacterized environments such as soil.

Analysis of tetranucleotide frequencies (TETRA) (42) clustered the 21 mVCs into three groups that were associated with the *Methylocystaceae*, *Methylococcoceae*, and an unknown group (Fig. 2*B*). The majority of viruses infected members of the *Methylocystaceae* family, with those infecting populations of the *Methylocystis* and *Methylosinus* genera restricted to acidic and neutral pH soils, respectively. TETRA correlation coefficients of all *Methylocystaceae*-linked viruses were in the same range both within and between either genus, suggesting coevolution with their host rather than genetic drift and divergence was the primary mechanism for defining specific associations with *Methylocystis* or *Methylosinus* strains.

Linkage of Active Viruses and Hosts from Analysis of Shared Homologs. Identification of the most abundant homologous genes between an individual mVC and one prokaryotic taxonomic family were always consistent with host–virus linkages established using spacer sequences from MAG CRISPR arrays. Specifically, BLASTp searches of genes present in the nine mVCs linked to *Methylocystaceae* MAGs via CRISPR spacer sequences all contained "best hits" (amino acid identity >30%, e-value <10^{-5}, bit score >50, and query coverage >70%) to a minimum of five homologs also found in *Methylocystaceae* genomes. This was therefore used as a further criterion for establishing host–virus linkages. Analysis of assembled contigs from 12 metagenomic libraries of total community or

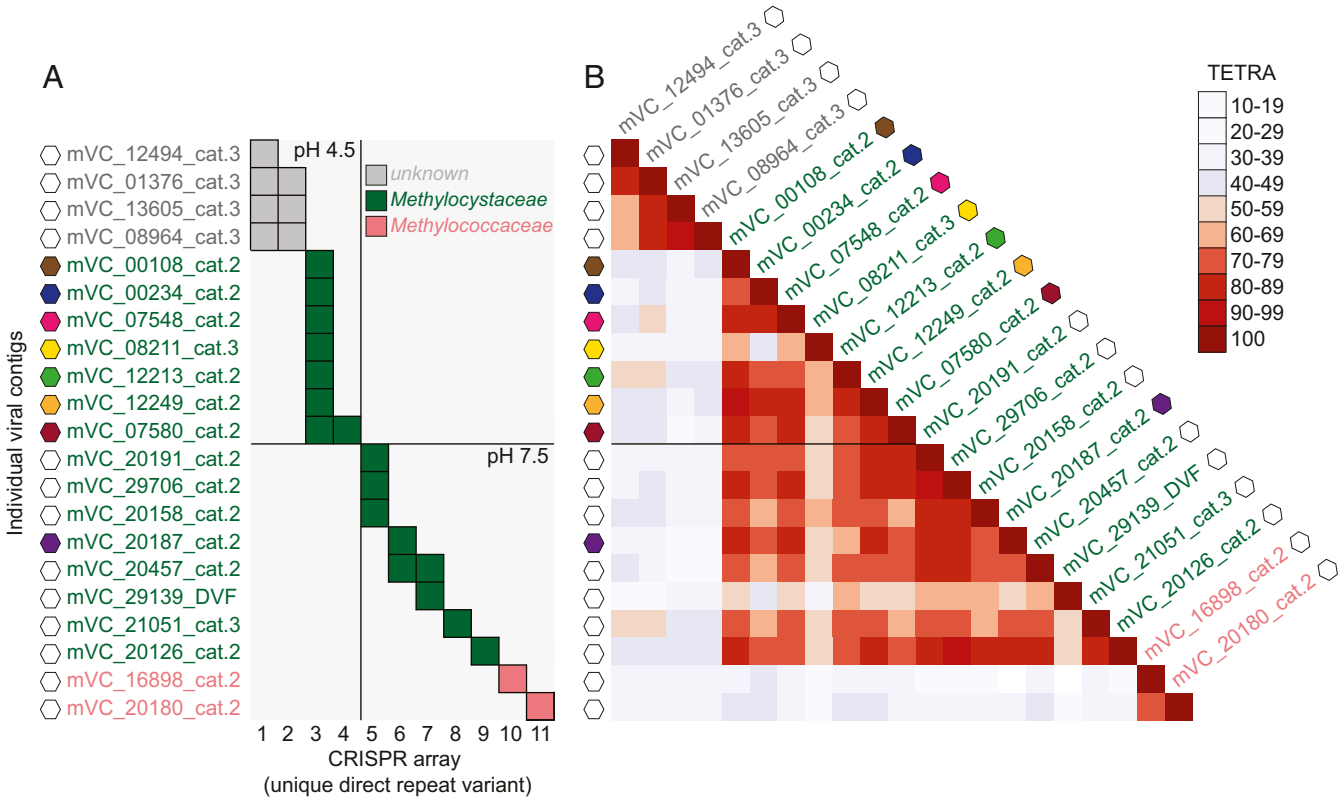

Fig. 2. Linkages of ^{13}C-enriched viruses to CRISPR arrays in pH 4.5 and 7.5 soil. (*A*) Matching protospacer sequences in 21 mVCs (hexagon symbols) with spacer sequences in 11 different CRISPR array variants (unique DR sequence) with 100% identity. Taxonomic affiliation of CRISPR arrays to host families was determined by phylogenomic analysis of affiliated MAGs (3, 6) or unbinned contigs (4, 5, 7–9), or inferred from shared homologs between linked mVCs and bacterial genomes (10, 11). mVC names also describe prediction by VirSorter (cat.2 or cat.3) or by DeepVirFinder alone (DVF). All mVCs were >10 kb except mVC_08964_cat.3 (9.8 kb) and mVC_28139_DVF (5.1 kb). These mVCs were also the only two predicted using DVF, with calculated probabilities describing "likely" and "probable" viruses, respectively. (*B*) TETRA correlation coefficients between 21 CRISPR-linked mVCs. Color-coded hexagon symbols denote mVCs linked to CRISPR arrays in Fig. 1.

virome DNA from the same soil samples without CH_4 incubation (Dataset S1) contained only three VirSorter-predicted mVCs that were linked to methanotrophs. In contrast, using a ^{13}C-targeted approach, 63% of mVCs contained a homolog that was linked to genomes of known C1-utilizing bacteria, with 35% linked specifically to populations from the *Methylocystaceae*, *Methylococcaceae*, or *Hyphomicrobiaceae* (Fig. 3*A*). While analysis of bacterial homologs in mVCs identified the taxonomic family of the assumed dominant host, they also indicated that individual viruses may infect hosts of other families of the same taxonomic order, including those at other trophic levels. Specifically, within the *Rhizobiales*, mVCs linked to *Methylocystaceae* also contained homologs shared with *Bradyrhizobiaceae*, *Methylobacteriaceae*, and *Rhizobiaceae* (Fig. 3*B*), indicating that viruses of methanotrophs may also infect nonmethanotrophic methylotrophs that are active at the same time.

Methane-Derived Carbon in Viruses of Hosts from Different Trophic Levels. CH_4-derived C was also transferred to viruses of secondary (and potentially tertiary) utilizers. One group of mVCs were linked to methylotrophic *Hyphomicrobiaceae* and a second to a phylogenetically diverse range of nitrogen-fixing Rhizobia (i.e., *Bradyrhizobiaceae*, *Phyllobacteriaceae*, and *Rhizobiaceae*). These lineages contain known methylotrophs, methanol dehydrogenases have been identified in a range of rhizobial species (43), and these mVCs also contained homologs found in the genomes of nodulating *Methylobacterium* strains (44). Viruses of predatory *Bdellovibrio* and *Myxococcales* bacteria were predicted, consistent with the recovery of corresponding bacterial MAGs in ^{13}C-enriched DNA. One mVC (20210_cat.2) was linked to the genus *Bdellovibrio*, and three were linked to *Myxococcales* populations, containing gene homologous to four families within the order. While the latter were category-3 mVCs (i.e., possible viruses), *Myxococcales*-associated contigs were also predicted as viral in origin using other tools (45, 46). The high isotopic labeling of both heterotrophic predators (with no identifiable C1-utilizing capability) and their viruses indicate that the predators were feeding primarily on populations that incorporated CH_4-derived carbon (i.e., methanotrophs, nonmethanotrophic methylotrophs, or other organisms consuming metabolic products or biomass) as feeding on unlabeled bacteria would dilute their enrichment. While the path of carbon flow to predatory bacteria is uncertain, it indicates that they have a preference for preying upon growing populations rather than the majority not incorporating CH_4-derived C (27).

Gene-sharing network analysis of mVCs with viruses in the National Center for Biotechnology Information's Reference Sequence Database (NCBI RefSeq) and other metagenome studies

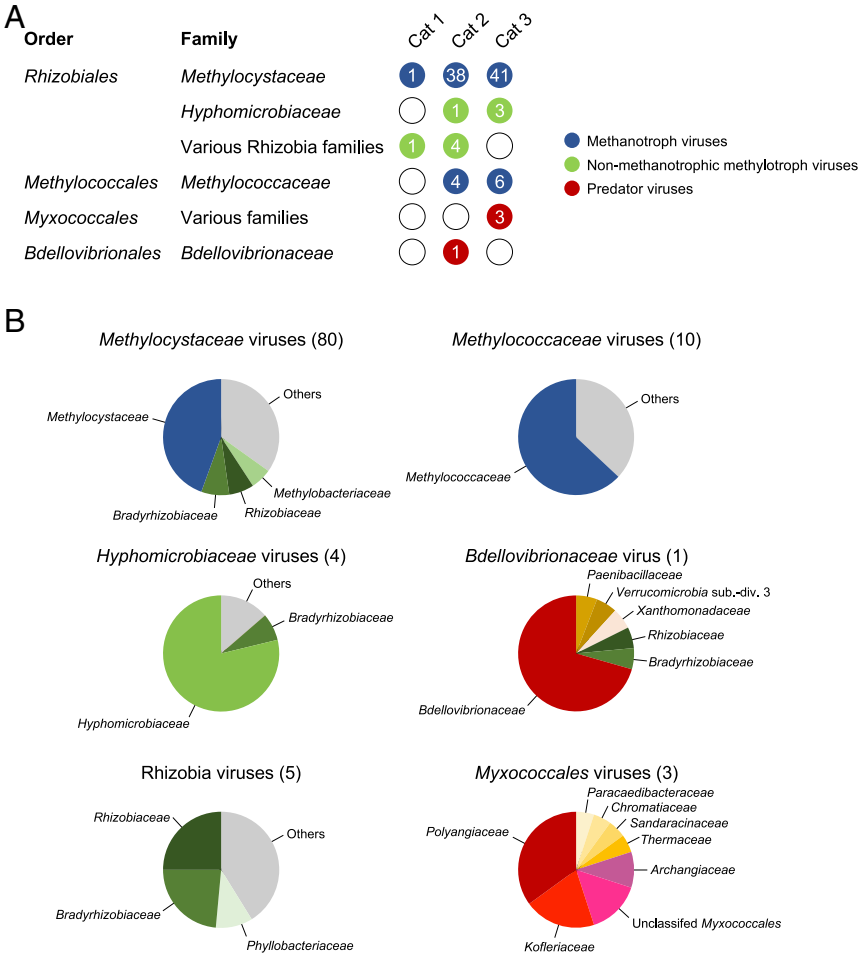

Fig. 3. Linkage of ^{13}C-enriched viruses to methanotrophic, nonmethanotrophic methylotrophic, and predator bacterial host populations through identification of shared homologous genes. (*A*) Association of viruses with different bacterial families and functional groups inferred from the presence of ≥5 shared homologous genes, with number of category-1, -2, and -3 VirSorter-predicted mVCs given. (*B*) Proportion of homologs in viruses linked to individual bacterial families. Each chart summarizes those mVCs that all contain ≥5 homologs to one family (number of mVCs given in parentheses) but with other taxonomic linkages also given. "Other" describes the proportion found in families each represented by less than <5% of homologs or those not annotated to the family level.

were analyzed using vConTACT 2.0 (47). Any linkages with RefSeq viruses typically had low scores (i.e., sharing a low number of homologs) and were linked to viruses of hosts that were inconsistent with our homolog-based predictions (Dataset S4). No linkages were observed with recently reported giant viruses of methanotrophs in freshwater lakes (22). However, in a recent study of 197 metagenomes from Swedish peatland soil, Emerson et al. (7) identified 13 viruses linked to methanotrophs. Intriguingly, eight of these were linked in our viral gene-sharing network, with both studies predicting *Methylocystaceae* hosts using different approaches for linkage prediction (*SI Appendix*, Fig. S4) and revealing the distribution of specific *Methylocystaceae*-associated viral groups present in different geographical areas and soil types. Analysis of gene-sharing networks of mVCs from this study indicated that there were two distinct *Methylocystaceae*-linked viral clusters which also varied in their distribution in both soils. Specifically, one cluster was associated with low pH only, whereas the second cluster contained viruses found in both pH 4.5 and 7.5 soils, including those linked by CRISPR spacer sequences. Individual networks of *Methylococcaceae*- and rhizobia-associated mVCs were also identified, associated with one of the two soils of contrasting pH. Taxonomically linked mVCs with ≥5 homologs were consistently placed in networks with other mVCs containing 1 to 4 homologs from the same methylotrophic families, confirming host linkage to a larger number of mVCs.

Genomic Content of ^{13}C-Enriched Viral Contigs. mVCs contained 8,174 genes, with 49.6% (4,054) annotated and representing 606 unique functions. Of these, genes encoding viral proteins accounted for 9.8% (397 genes) and included major capsid proteins, tail proteins, integrases, portal proteins, and terminases. Bacterial proteins used for viral replication accounted for 5.1% (206 genes). A number of metagenomic studies have demonstrated that viruses can possess genes encoding subunit C of ammonia or particulate methane monooxygenases as auxiliary metabolic genes (22, 48), which are also typically found as isolated genes in genomes in addition to being present in clusters or operons encoding A and B subunits (49). In this study, one low-confidence mVC (7.3 kb, category-3) was identified as containing an isolated *pmoC* gene that was phylogenetically related to growing *Methylocystis* populations but was distinct from *pmoC* sequences found in viruses associated with freshwater *Methylocystis* populations (22) (*SI Appendix*, Fig. S5).

Conclusions

These results demonstrate that by following carbon flow, native virus–host interactions associated with a critical biogeochemical process can be identified at the resolution of individual populations. Active interactions were observed at different trophic levels within the highly complex soil environment, from primary utilizers of CH_4 to heterotrophic bacteria preying upon ^{13}C-enriched methylotrophs or other organisms consuming CH_4-derived metabolic products or biomass. Therefore, viral lysis may be an important mechanism for the transfer of CH_4-derived organic carbon into a soil viral shunt. Type I and II methanotrophs interact with distinct groups of viruses, and the composition of CRISPR arrays of *Methylocystaceae* reveal that they have a continual dynamic interaction with specific viral populations. Analysis of shared homologs in individual viral genomes show that they may also interact with host populations at different trophic levels within a CH_4-fueled network.

Methods

Soil Microcosms. Triplicate soil samples were collected in February 2018 at 1-m intervals from the upper 10 cm of pH 4.5 and 7.5 soil subplots of a pH gradient maintained since 1961 and under an 8-y crop rotation (Woodlands Field, Craibstone Estate, Scotland's Rural College, Aberdeen, grid reference NJ872104) (50). The crop at the time of sampling was potatoes. Soil (podzol, sandy-loam texture) was sieved (2-mm mesh size) and microcosms established in triplicate for each soil pH and isotope in 144-mL serum bottles containing 14.30 g soil (10 g dry weight equivalent) with a 30% volumetric water content, equivalent to ∼60% water-filled pore space. Bottles were capped and established with a 10% (vol/vol) ^{12}C-CH_4 or ^{13}C-CH_4 (Sigma-Aldrich) headspace (99% atom enriched), reopening every 10 d to maintain aerobic conditions before sealing and re-establishing CH_4 headspace concentrations. Microcosms were incubated at 25 °C and destructively sampled after 30 d with soil archived immediately at −20 °C.

DNA Extraction and Stable-Isotope Probing. Genomic DNA was extracted from 0.5-g soil samples using a CTAB buffer phenol:chloroform: isoamyl alcohol bead-beating protocol and subjected to isopycnic centrifugation in CsCl gradients, recovery, and purification as previously described (51). Briefly, 6 μg genomic DNA was added to 8 mL CsCl-Tris EDTA solution (refractive index (RI) of 1.4010; buoyant density of 1.71 g · mL^{-1}) in polyallomer tubes before sealing and ultracentrifugation at 152,000 × *g* (50,000 rpm) in a MLN80 rotor (Beckman-Coulter) for 72 h at 25 °C. CsCl gradients were fractionated into 350-μL aliquots using an in-house semiautomated fraction recovery system before determining RI and recovering DNA. The relative abundance of bacterial 16S rRNA and methanotrophic *pmoA* genes in genomic DNA distributed across the CsCl gradients was determined by qPCR in a Corbett Rotor-Gene 6000 thermocycler (Qiagen) using primer sets P1(341f)/P2(534r) (52) and A189F/A682R (53), respectively. Reactions (25-μL) contained 12.5 μL 2X QuantiFast SYBR Green Mix (Qiagen), 1 μM each primer, 100 ng T4 gene protein 32 (Thermo Fisher), 2 μL standard (10^8 to 10^2 copies of an amplicon-derived standard), or 1/10 diluted DNA. Thermocycling conditions consisted of an initial denaturation step of 15 min at 95 °C for both assays followed by 30 cycles of 15 s at 94 °C, 30 s at 60 °C, and 30 s at 72 °C for the 16S rRNA gene assay or 60 s at 94 °C, 60 s at 56 °C, and 60 s at 72 °C for the *pmoA* gene assay, followed by melt-curve analysis. All assays had an efficiency between 93 and 97% with an r^2 value >0.99. Genomic DNA from four fractions with a buoyant density >1.732 g · mL^{-1} were then pooled for each ^{12}C- and ^{13}C-CH_4–incubated replicate for 16S rRNA gene amplicon sequence and metagenomic analysis.

DNA extraction for soil virome analysis is summarized in *SI Appendix*.

Metagenome Sequencing, Assembly, and Annotation. Library preparation and sequencing was performed at the Joint Genome Institute (JGI). Libraries were produced from fragmented DNA using KAPA Biosystems Library Preparation Kits (Roche) and quantified using KAPA Biosystems NGS library qPCR kits. Indexed samples were sequenced (2 × 150 bp) on the Illumina NovaSeq platform with NovaSeq XP version 1 reagent kits and a S4 flowcell. Raw reads were processed with JGI's RQCFilter2 pipeline that utilized BBTools version 38.51 (54). Reads containing adapter sequences were trimmed, and those with ≥3 N bases or ≤51 bp or ≤33% of full-read length were removed along with PhiX sequences using BBDuk, and reads mapped to human, cat, dog, or mouse references at 95% identity were removed using BBMap. De novo contig assembly of the 100 to 196 million quality-controlled reads per metagenome was performed using MetaSPAdes version 3.13.0 (55). The 1 to 2 million contigs per metagenome were then concatenated together, and contigs larger than 5 kb were dereplicated at 99% ANI using PSI-CD-HIT version 4.6.1 (56) and binned using MetaWRAP version 1.2.1 (57) (Dataset S5). Bin completion and contamination was determined by CheckM version 1.0.12 (58). Taxonomic annotation of contigs was performed using Kaiju (59) with the NCBI RefSeq database (Release 94; 25 June 2019) (60) and MAGs using GTDB-Tk version 0.3.2 (61) with the Genome Taxonomy Database (release 89, 21 June 2019) (62). Protein sequence annotation was performed using InterProScan 5 (e-value <10^{-5}) (63). Pairwise ANI comparison of MAGs was calculated using FastANI (39).

Amplicon Sequencing and Analysis. 16S rRNA genes were amplified using primers 515F/806R (64) followed by library preparation and sequencing on an Illumina MiSeq sequencer as previously described (65). Reads with a quality score <20 and length <100 bp were discarded using FASTX-Toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). High-quality reads were merged using PANDAseq version 2.11 (66) and denoising and chimera removal performed with UNOISE3 (67). Amplicon sequence variants (ASVs) were annotated using the RDP classifier version 2.11 (68). Nonmetric multidimensional scaling of Bray–Curtis dissimilarity derived from the relative abundance of ASVs was performed with the metaMDS function in the vegan package (69) in R version 3.6.0.

Virus Prediction. mVCs were predicted from 9,190 contigs ≥10 kb using VirSorter (32), retaining nonprophage category-1, -2, or -3 mVCs, representing "most

confident," "likely," and "possible." DeepVirFinder (33) was also used to predict mVCs from contigs ≥10 kb, with those with a *P* value <0.05 and a score ≥0.9 or 0.7, representing "confident" and "possible," respectively (Dataset S6). mVCs predicted from shorter contigs (≥5 and <10 kb) were used for identifying CRISPR spacer sequence matches only. The relative abundance of each mVC in the six metagenomes was determined using the MetaWRAP-Quant_bins module (57) and a heatmap produced using the heatmaply package in R version 3.6.0. The tools CheckV (45) and VIBRANT (46) were also used to predict a viral origin of category-3 *Myxococcales*-associated mVCs.

Virus–Host Linkage. CRISPR arrays within MAGs and unbinned contigs were identified using the CRISPR Recognition Tool version 1.2 (37) (Dataset S7). DR and spacer sequences were extracted before performing searches against positive and negative strands to identify MAGs or contigs with direct repeats and the viral origin of spacers using Seqkit commands (38). After identification of matched spacer sequences in mVCs, 10 nucleotides before and after the spacer sequence were extracted to identify associated host-specific PAM sequences. Conserved and variant PAM sequences were manually identified. Correlation coefficients of pairwise comparison of the tetranucleotide frequencies between unique CRISPR-associated mVCs were calculated using Python package pyani version 0.2.10 (70). To identify homologous genes shared between CRISPR-linked viruses and hosts, gene prediction was performed using Prodigal version 2.6.3 (71) with the -p meta option followed by protein alignment with Blastp (identity >30%, e-value $<10^{-5}$, bit score >50, and query cover >70%) and protein sequence annotation using InterProScan 5 (e-value $<10^{-5}$). Gene homology between all mVCs and prokaryotes in the NCBI nr database was determined using Diamond Blastp (e-value $<10^{-5}$) (72). Virus–host prediction using *k*-mer frequencies was performed with WIsH version 1.0 (73). Networks based on shared gene content was constructed using vConTACT 2.0 (47) with the NCBI RefSeq database (Release 94; 25 June 2019).

Phylogenetic Analysis of PmoC and PxmC Protein Sequences. Maximum likelihood analysis of inferred protein sequences of membrane-bound monooxygenase C subunits from methanotroph MAGs and reference sequences (Dataset S8) was performed on 229 unambiguously aligned sequences using PhyML 3.0 (74) with automatic model selection (LG substitution, gamma distribution [0.06], and proportion of invariable sites [0.087] estimated). Bootstrap support was calculated from 100 replicates.

Data Availability. Metagenome sequence reads are deposited in NCBI's GenBank under BioProject accession nos. PRJNA621430–PRJNA621447. Metagenome draft assemblies are accessible through the JGI Genome Portal (DOI: 10.25585/1487501). Amplicon sequence data are deposited in the NCBI Sequence Read Archive with BioProject accession no. PRJNA676099.

ACKNOWLEDGMENTS. This work was funded by an AXA Research Chair awarded to G.W.N., a France-Berkeley Fund grant (2018 to 2019) awarded to G.W.N. and M.K.F., and the US Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research Genomic Science Program under Award No. DE-SC0010570 to M.K.F. The sequencing data were generated under JGI Community Science Program Proposal No. 503702 awarded to G.W.N. and C.H. The work conducted by the US DOE JGI, a DOE Office of Science User Facility, is supported by the Office of Science of the US DOE under Contract No. DE-AC02-05CH11231. The pH gradient experiment is funded through the Scottish Government's Rural and Environment Science and Analytical Services 2016 to 2021 program. We would like to thank Prof. Joanne Emerson for valuable discussion and Dr. Laurent Pouilloux for assistance with the Newton high-performance computing cluster at École Centrale de Lyon.

1. A. Frossard, F. Hammes, M. O. Gessner, Flow cytometric assessment of bacterial abundance in soils, sediments and sludge. *Front. Microbiol.* **7**, 903 (2016).
2. K. E. Williamson, J. J. Fuhrmann, K. E. Wommack, M. Radosevich, Viruses in soil ecosystems: An unknown quantity within an unexplored territory. *Annu. Rev. Virol.* **4**, 201–219 (2017).
3. C. A. Suttle, Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
4. J. B. Emerson, Soil viruses: A new hope. *mSystems* **4**, e00120-19 (2019).
5. J. C. Ignacio-Espinoza, N. A. Ahlgren, J. A. Fuhrman, Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nat. Microbiol.* **5**, 265–271 (2020).
6. P. Gómez, A. Buckling, Bacteria-phage antagonistic coevolution in soil. *Science* **332**, 106–109 (2011).
7. J. B. Emerson *et al*., Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).
8. P. Carini *et al*., Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2**, 16242 (2016).
9. M. Saunois *et al*., The global methane budget 2000–2017. *Earth Syst. Sci. Data* **12**, 1561–1623 (2020).
10. M. Etminan, G. Myhre, E. J. Highwood, K. P. Shine, Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.* **12**, 614–623 (2016).
11. J. Le Mer, P. Roger, Production, oxidation, emission and consumption of methane by soils: A review. *Eur. J. Soil Biol.* **37**, 25–50 (2001).
12. R. Angel, P. Claus, R. Conrad, Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J.* **6**, 847–862 (2012).
13. C. Knief, Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Front. Microbiol.* **6**, 1346 (2015).
14. Z. Lyu, N. Shao, T. Akinyemi, W. B. Whitman, Methanogenesis. *Curr. Biol.* **28**, R727–R732 (2018).
15. S. A. Morris, S. Radajewski, T. W. Willison, J. C. Murrell, Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing. *Appl. Environ. Microbiol.* **68**, 1446–1453 (2002).
16. J. Pratscher, J. Vollmers, S. Wiegand, M. G. Dumont, A.-K. Kaster, Unravelling the identity, metabolic potential and global biogeography of the atmospheric methane-oxidizing Upland Soil Cluster α. *Environ. Microbiol.* **20**, 1016–1029 (2018).
17. A. T. Tveit *et al*., Widespread soil bacterium that oxidizes atmospheric methane. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8515–8524 (2019).
18. J. Zhao, Y. Cai, Z. Jia, The pH-based ecological coherence of active canonical methanotrophs in paddy soils. *Biogeosciences* **17**, 1451–1462 (2020).
19. L. Chistoserdova, M. G. Kalyuzhnaya, M. E. Lidstrom, The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* **63**, 477–499 (2009).
20. F. M. Tyutikov, I. A. Bespalova, B. A. Rebentish, N. N. Aleksandrushkina, A. S. Krivisky, Bacteriophages of methanotrophic bacteria. *J. Bacteriol.* **144**, 375–381 (1980).
21. F. M. Tyutikov *et al*., Bacteriophages of methanotrophs isolated from fish. *Appl. Environ. Microbiol.* **46**, 917–924 (1983).
22. L.-X. Chen *et al*., Large freshwater phages with the potential to augment aerobic methane oxidation. *Nat. Microbiol.* **5**, 1504–1515 (2020).
23. S. Almumin, M. Kadri, U. Mamat, J. Engel, Isolation and primary characterization of a new bacteriophage of obligate methylotrophic bacteria. *J. Basic Microbiol.* **30**, 627–633 (1990).
24. H. H. Buchholz *et al*., Efficient dilution-to-extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria. *ISME J.* **15**, 1585–1598 (2021).
25. M. Yang *et al*., Genomic characterization and distribution pattern of a novel marine OM43 phage. *Front. Microbiol.* **12**, 651326 (2021).
26. S. Radajewski *et al*., Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. *Microbiology (Reading)* **148**, 2331–2342 (2002).
27. B. A. Hungate *et al*., The functional significance of bacterial predators. *MBio* **12**, e00466-21 (2021).
28. A. L. Pasulka *et al*., Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ. Microbiol.* **20**, 671–692 (2018).
29. R. M. Bowers *et al*., Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
30. J. Pérez, A. Moraleda-Muñoz, F. J. Marcos-Torres, J. Muñoz-Dorado, Bacterial predation: 75 years and counting! *Environ. Microbiol.* **18**, 766–779 (2016).
31. S. Roux *et al*., Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
32. S. Roux, F. Enault, B. L. Hurwitz, M. B. Sullivan, VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
33. J. Ren *et al*., Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).
34. D. Paez-Espino, G. A. Pavlopoulos, N. N. Ivanova, N. C. Kyrpides, Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. *Nat. Protoc.* **12**, 1673–1682 (2017).
35. R. Sausset, M. A. Petit, V. Gaboriau-Routhiau, M. De Paepe, New insights into intestinal phages. *Mucosal Immunol.* **13**, 205–215 (2020).
36. M. T. Jahn *et al*., Lifestyle of sponge symbiont phages by host prediction and correlative microscopy. *ISME J.* **15**, 2001–2011 (2021).
37. C. Bland *et al*., CRISPR recognition tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
38. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962 (2016).
39. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
40. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155**, 733–740 (2009).
41. R. T. Leenay, C. L. Beisel, Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.* **429**, 177–191 (2017).
42. J. Wang *et al*., Genomic sequence of *'Candidatus liberibacter solanacearum'* haplotype C and its comparison with haplotype A and B genomes. *PLoS One* **12**, e0171531 (2017).

43. J. Huang *et al*., Rare earth element alcohol dehydrogenases widely occur among globally distributed, numerically abundant and environmentally important microbes. *ISME J.* **13**, 2005–2017 (2019).
44. P. N. Green, J. K. Ardley, Review of the genus *Methylobacterium* and closely related organisms: A proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov. *Int. J. Syst. Evol. Microbiol.* **68**, 2727–2748 (2018).
45. S. Nayfach *et al*., CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
46. K. Kieft, Z. Zhou, K. Anantharaman, VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).
47. H. Bin Jang *et al*., Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
48. N. A. Ahlgren, C. A. Fuchsman, G. Rocap, J. A. Fuhrman, Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode *amoC* nitrification genes. *ISME J.* **13**, 618–631 (2019).
49. G. W. Nicol, C. Schleper, Ammonia-oxidising Crenarchaeota: Important players in the nitrogen cycle? *Trends Microbiol.* **14**, 207–212 (2006).
50. J. S. Kemp, E. Paterson, S. M. Gammack, M. S. Cresser, K. Killham, Leaching of genetically modified *Pseudomonas fluorescens* through organic soils: Influence of temperature, soil pH, and roots. *Biol. Fertil. Soils* **13**, 218–224 (1992).
51. G. W. Nicol, J. I. Prosser, Strategies to determine diversity, growth, and activity of ammonia-oxidizing archaea in soil. *Methods Enzymol.* **496**, 3–34 (2011).
52. G. Muyzer, E. C. de Waal, A. G. Uitterlinden, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**, 695–700 (1993).
53. A. J. Holmes, A. Costello, M. E. Lidstrom, J. C. Murrell, Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* **132**, 203–208 (1995).
54. B. Bushnell, BBTools software package. https://sourceforge.net/projects/bbmap/. Accessed 30 July 2021.
55. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
56. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
57. G. V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
58. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
59. P. Menzel, K. L. Ng, A. Krogh, Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
60. N. A. O'Leary *et al*., Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44** (D1), D733–D745 (2016).
61. P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
62. D. H. Parks *et al*., A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
63. P. Jones *et al*., InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
64. W. Walters *et al*., Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**, e00009–e00015 (2015).
65. D. R. Finn, S. Lee, M. B. Lazén, G. W. Nicol, C. Hazard, Cropping systems that improve richness convey greater resistance and resilience to soil fungal, relative to prokaryote, communities. *bioRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.03.15.992560. Accessed 30 July 2021.
66. A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, J. D. Neufeld, PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
67. R. C. Edgar, UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* [Preprint] (2016). https://doi.org/10.1101/081257. Accessed 30 July 2021.
68. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
69. J. Oksanen *et al*., vegan: Community Ecology Package. https://CRAN.R-project.org/package=vegan (2019). Accessed 30 July 2021.
70. L. Pritchard, R. H. Glover, S. Humphris, J. G. Elphinstone, I. K. Toth, Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* **8**, 12–24 (2016).
71. D. Hyatt *et al*., Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
72. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
73. C. Galiez, M. Siebert, F. Enault, J. Vincent, J. Söding, WIsH: Who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114 (2017).
74. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).